BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

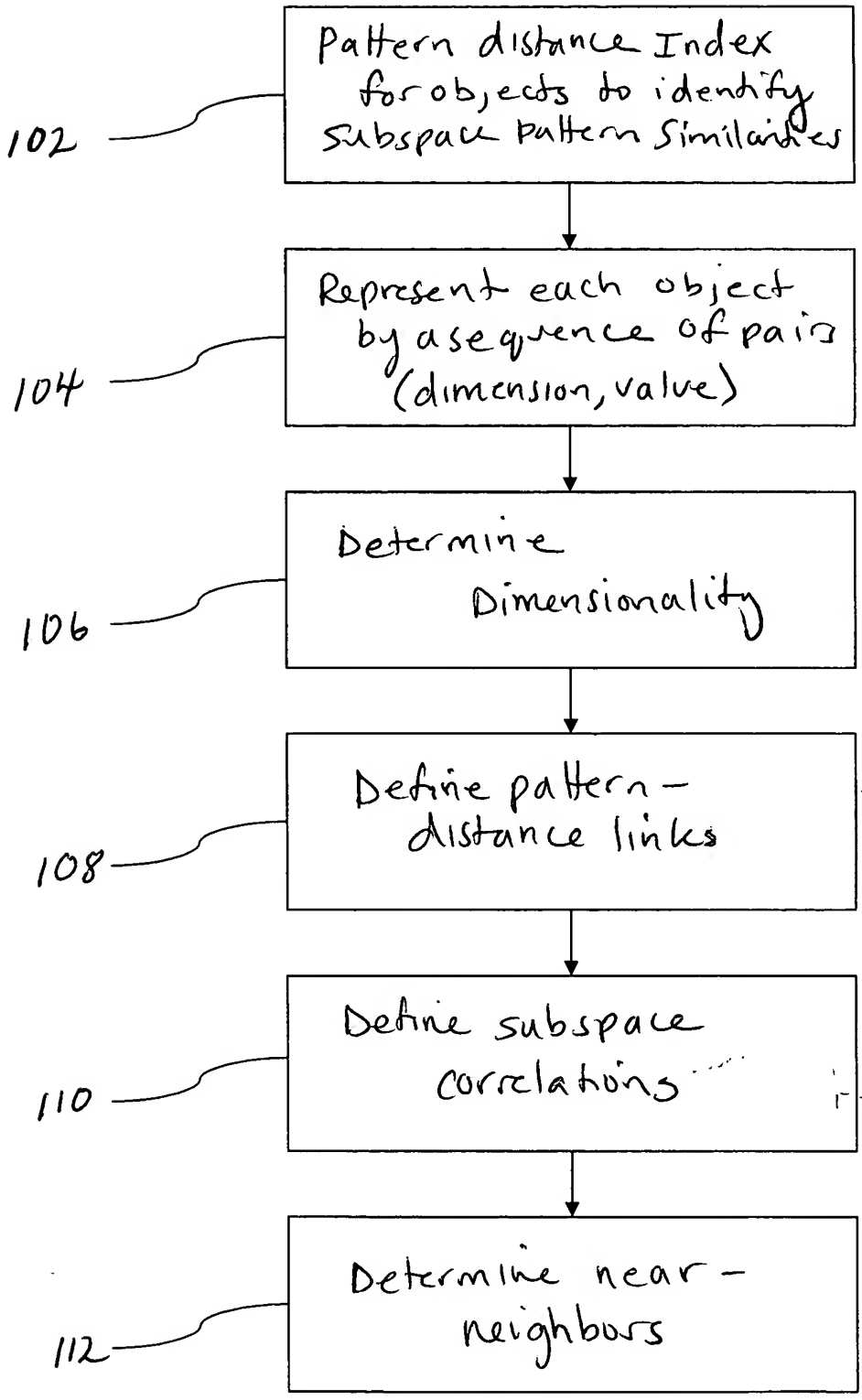

Pattern distance Index for objects to identify subspace pattern similarities
102
Represent each object by a sequence of pairs (dimension, value)
104
Determine Dimensionality
106
Define pattern-distance links
108
Define subspace correlations
110
Determine near-neighbors
112

FIG. 1

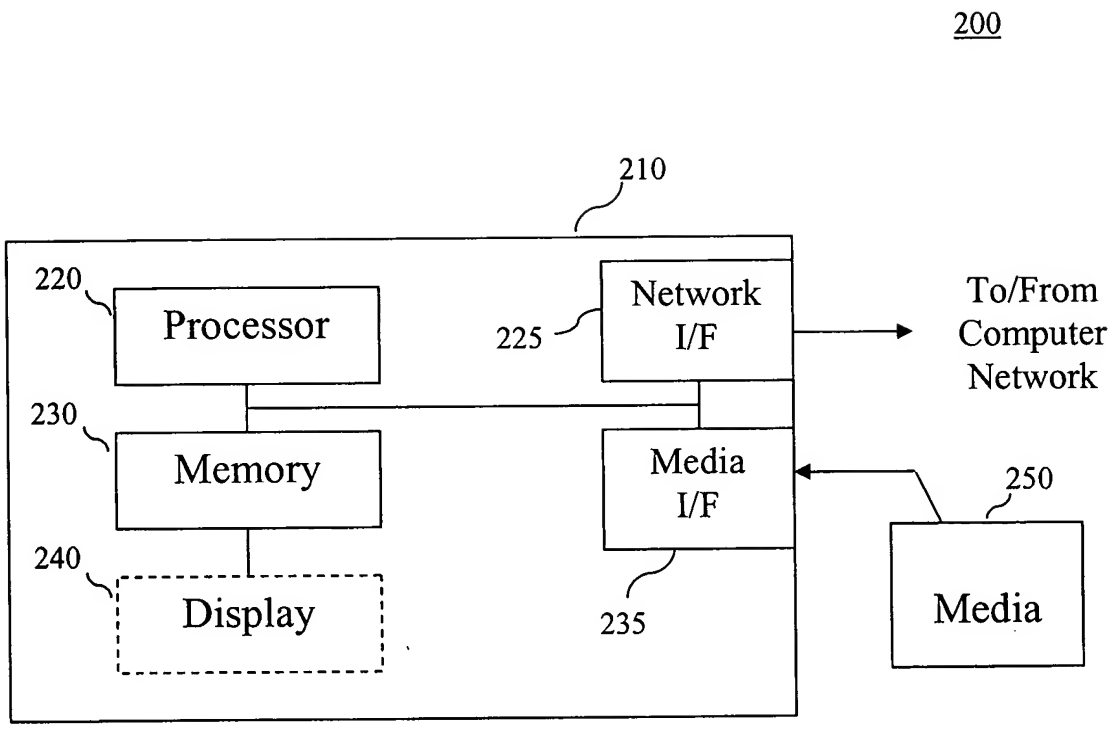

200
210
220
Processor
225
Network
I/F
To/From
Computer
Network
230
Memory
Media
I/F
250
240
Display
235
Media

FIG. 2

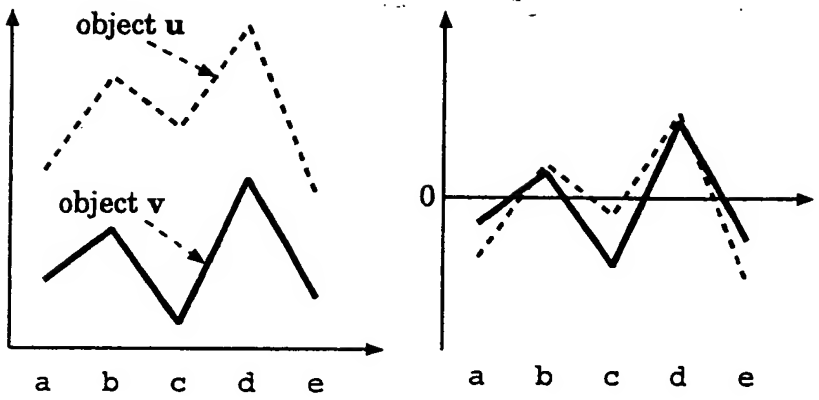

object u
object v
a b c d e
0
a b c d e

FIG. 3

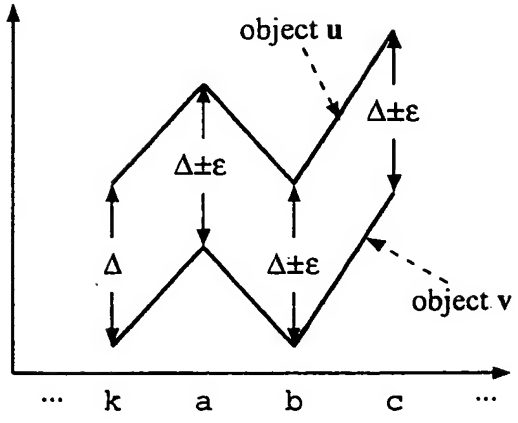

object u
Δ±ε
Δ±ε
Δ
Δ±ε
object v
··· k a b c ···

FIG. 4

$f(u,i)$, where $u \in \{\#1, \#2\}$ and $i = 1, \cdots, 4$				
(c_1, 0),	(c_2, -3),	(c_3, 1),	(c_4, -1),	(c_5, -3)
	(c_2, 0),	(c_3, 4),	(c_4, 2),	(c_5, 0)
		(c_3, 0),	(c_4, -2),	(c_5, -4)
			(c_4, 0),	(c_5, -2)
(c_1, 0),	(c_2, -3),	(c_3, 1),	(c_4, -1),	(c_5, 2)
	(c_2, 0),	(c_3, 4),	(c_4, 2),	(c_5, 5)
		(c_3, 0),	(c_4, -2),	(c_5, 1)
			(c_4, 0),	(c_5, 3)

FIG. 5

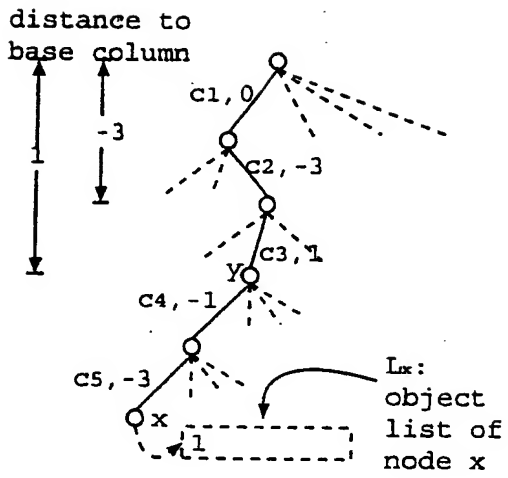

distance to
base column
-3
1
c1, 0
c2, -3
c3, 1
y
c4, -1
c5, -3
x
1
Lx:
object
list of
node x

FIG. 6

Input: T: a trie built on $\mathcal{D}$
S: a subspace defined by a continuous column set $\{c_i, c_{i+1}, ..., c_k\}$
$q = (c_1, v_1), \cdots, (c_n, v_n)$: a query object
ϵ: pattern threshold

Output: near-neighbors of q in subspace S

$n \leftarrow$ root of T;
$search(n, S)$;

Function $search(x, S)$
if $S = \emptyset$ **then**
    output the descendents of x;
else
    assume $S = \{c_j, c_{j+1}, ..., c_k\}$;
    for *x's child node y under edge labeled* (c_j, v) *where* $v \in [(v_j - v_i) - \epsilon, (v_j - v_i) + \epsilon]$ **do**
        $search(y, \{c_{j+1}, ..., c_k\})$;

FIG. 7

Input: $\mathcal{D}$: objects in multi-dimensional space $\mathcal{A}$
Output: PD-Index of $\mathcal{D}$

for *each* $u \in \mathcal{D}$ **do**
  insert $f(u, i), 1 \leq i < |\mathcal{A}|$ into a trie; (Eq 5)

for *each node* x *encountered in a depth-first traversal of the trie* **do**
  label node x by $\langle n_x, s_x \rangle$;
  let (c, d) be the arc that points to x;
  append $\langle n_x, s_x \rangle$ to pattern-distance link (c, d);

FIG 8

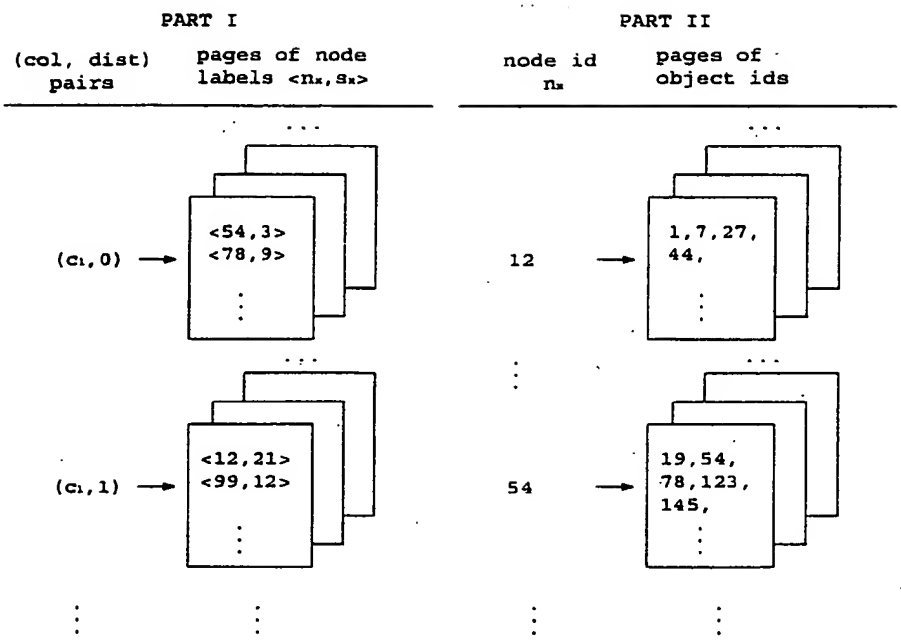

PART I
(col, dist) pairs
pages of node labels <nx, sx>
(c1, 0)
<54,3>
<78,9>
(c1, 1)
<12,21>
<99,12>
PART II
node id nx
pages of object ids
12
1,7,27,
44,
54
19,54,
78,123,
145,

FIG. 9A

FIG. 9B

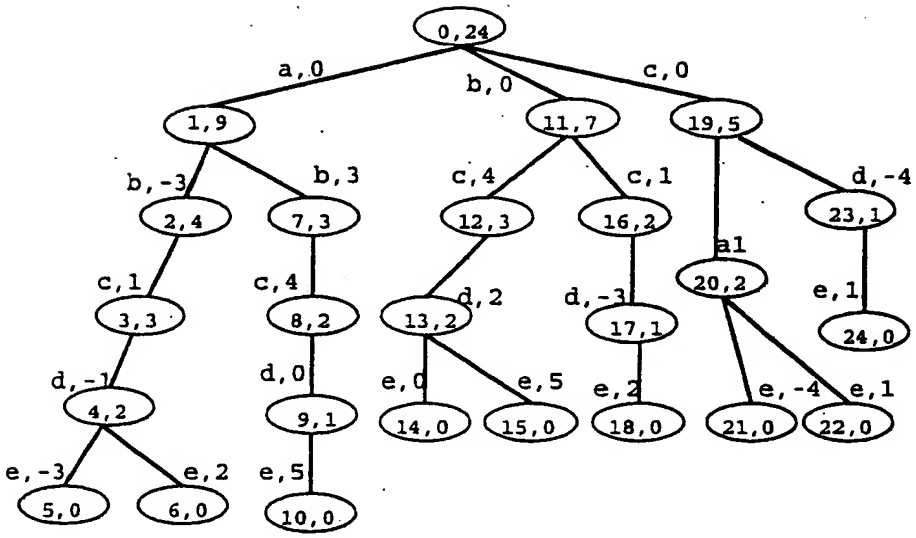
Input: q: a query object, S: a given subspace
ϵ: pattern threshold

Output: q's near-neighbors in subspace S

let $(c_1, v_1), \cdots, (c_{|S|}, v_{|S|})$ be q's projection on S;
$x \leftarrow$ the node under arc $(c_1, 0)$;
$search(x, 2)$;

Function $search(x, i)$
if $i \leq |S|$ **then**
  for *pattern distance link* I *of* (c_i, v), *where* $v \in [v_i - v_1 - \epsilon, v_i - v_1 + \epsilon]$ **do**
    /* *perform a binary search on* I */
    for *all node* $r \in I$ *and* $n_r \in [n_x, n_x + s_x]$ **do**
      $search(r, i + 1)$;
    end
  end
else
  output objects in L_x, $x = v_s, ..., v_m$
end

FIG. 10



node	5	6	10	14	15	18	21	22	24
objs	{1}	{2}	{3,4}	{1}	{2}	{3,4}	{1}	{2}	{3,4}

FIG. 11

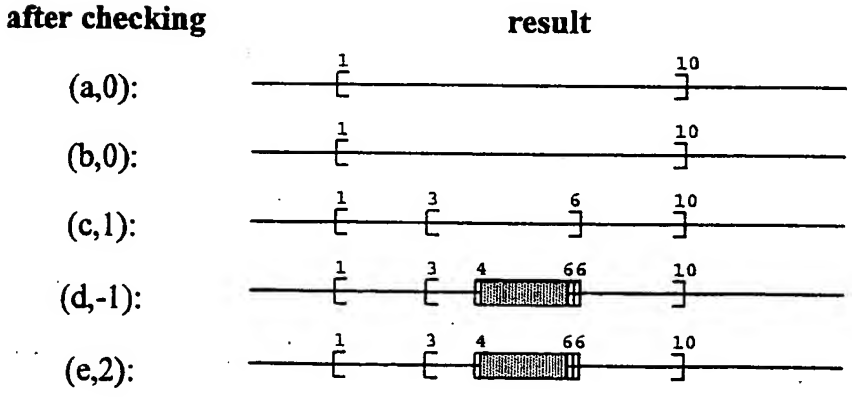
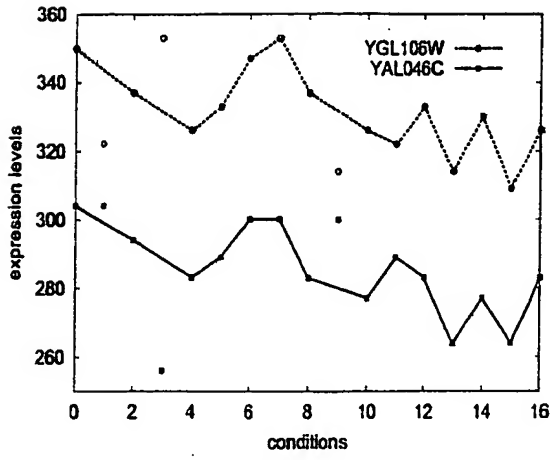


FIG. 12

Input: $q = (c_1, v_1), \cdots, (c_n, v_n)$: a query object
r: distance threshold, ϵ: pattern tolerance
F: index file for $\mathcal{D}$
Output: $\mathcal{NN}(q, r)$

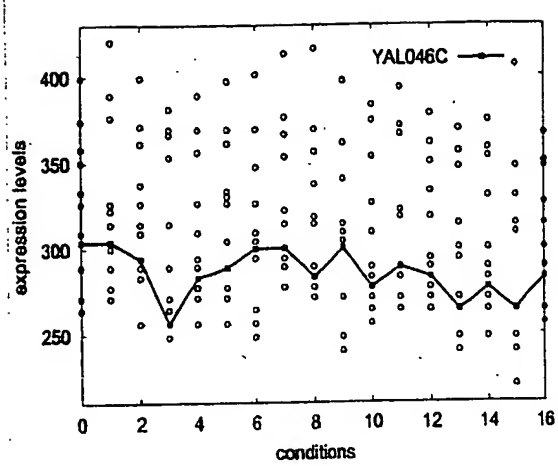
for $i = 1, ..., r + 1$ **do**
  $R \leftarrow$ the range of the (only) node in link $(c_i, 0)$;
  $j \leftarrow i + 1$;
  while $R \neq \phi$ **and** $j \leq |\mathcal{A}|$ **do**
    search link (c_j, v) for nodes inside any range of R, where $v \in [v_j - v_i - \epsilon, v_j - v_i + \epsilon]$;
    update R by adding the ranges of those nodes;
    if *a region s of R is inside $|\mathcal{A}| - r$ brackets* **then**
      output objects in L_x where $x \in s$;
      eliminate s from R;
    end
    if *a region s of R is inside less than $r - j$ brackets* **then**
      eliminate the region from s;
    end
    $j \leftarrow j + 1$;
  end
end

FIG. 13



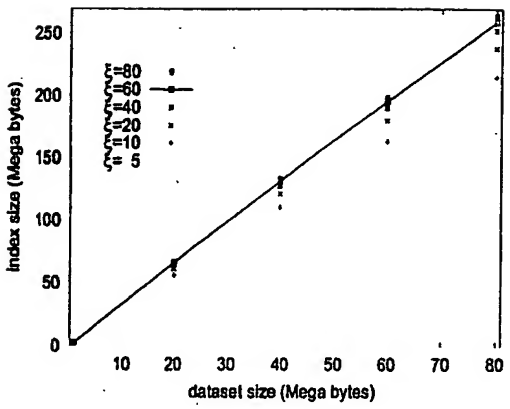
$pdist(\cdot,\cdot) \leq 3$

FIG. 14A



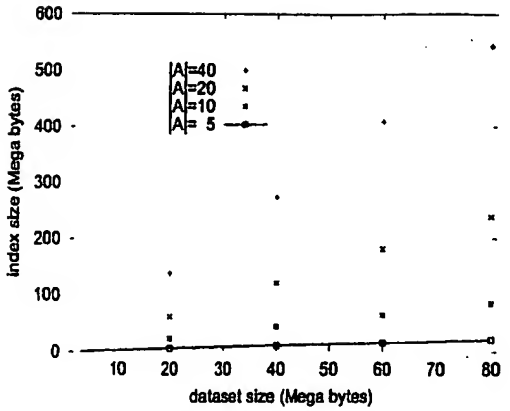
$pdist(\cdot,\cdot) \leq 4$

FIG. 14B



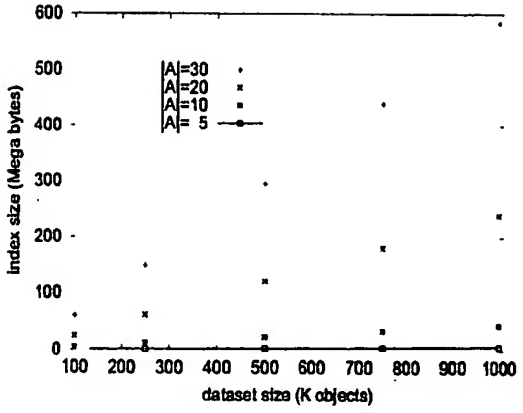
$|A| = 20, \xi = 5, ..., 80$

FIG. 15A



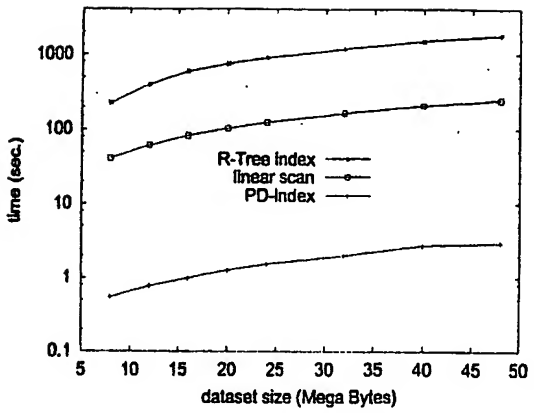
varying total data size, $\xi = 20$

FIG. 15B



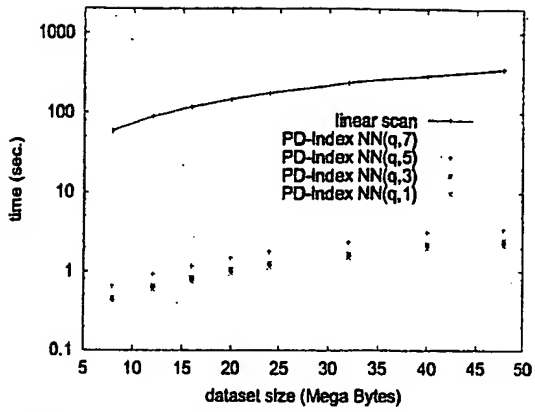
varying # of objects, $\xi = 20$

FIG. 15C



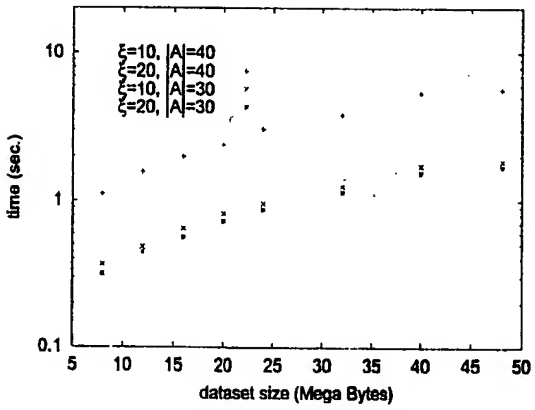
Pattern matching in given subspaces

FIG. 16A



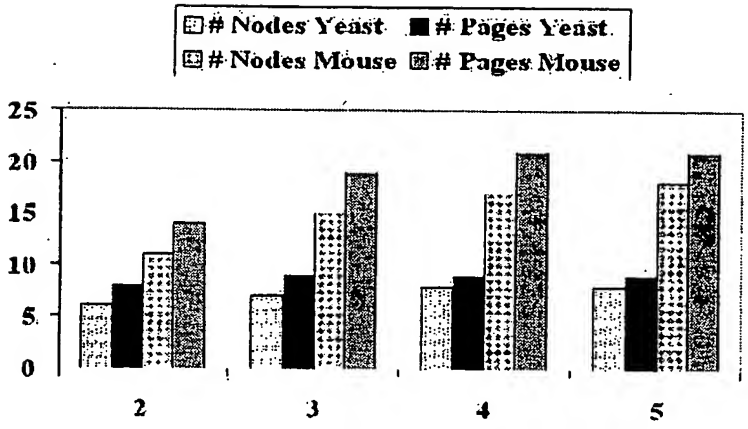
Near-neighbor search in subspaces beyond given dimensionalities

FIG. 16B



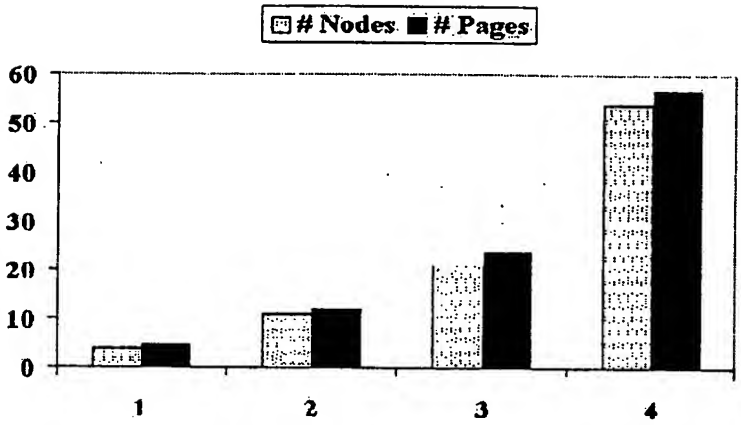
Impact of ξ and $|\mathcal{A}|$ in near-neighbor query $\mathcal{NN}(q, 7)$

FIG. 16C



(a) Find Near-neighbors in DNA micro-array in given subspaces (X axis is query length)

FIG. 17A



(b) Find Near-Neighbors in DNA micro-array (X axis is the distance radius r)

FIG. 17B